DOCUMENT RESUME

ED 146 136 SP 011 776

AUTHOR Masters, James R.
TITLE Improving Instruction Through Student Observation of Teaching Methods.
INSTITUTION Pennsylvania State Dept. of Education, Harrisburg. Bureau of Information Systems.
PUB DATE Sep 77
NOTE 44p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Questionnaires; *Secondary Education; *Student Developed Materials; *Student Evaluation of Teacher Performance; Teacher Evaluation; Teaching Skills; Test Reliability; Test Validity
IDENTIFIERS Student Observation Teachers Teaching Techniques

ABSTRACT

This study was undertaken to examine the reliability and validity of a student-designed instrument developed to provide feedback to high school teachers on student perception of teacher performance. The Student Observation of Teachers and Teaching Techniques (StOTT) questionnaire was developed by a board of eleventh- and twelfth-grade Pennsylvania students. The questionnaire contains twenty-nine positive statements concerning a teacher's classroom behavior, with a five-point scale provided for student responses. The statements are arranged in five subsets: Student-Teacher Relations, Grades and Testing, Materials, Teacher Personality, and Teaching Methods and Techniques. Three final questions investigate: (1) distracting personal mannerisms; (2) reasons for taking the course; and (3) feelings about recommending the course to others. Testing of the instrument was performed with a group of thirty-six teachers and 925 students divided into three treatment groups. Analyses of student responses and comparison of these analyses with previous research on teacher rating by students suggests that the StOTT is highly useful. Students seemed capable of making and repeating stable ratings over a period of time among the items. The subscales of the instrument make sense conceptually and in general are internally consistent. The major problem uncovered is that students of differing achievement levels will tend to rate teachers differently, with the more successful students giving higher StOTT ratings, and those who are "turned off" to school giving lower ratings. The major conclusion of the study is that, while the StOTT should not be used to evaluate or compare high school teachers, it is a valuable instrument for teachers to use in their own classrooms. (MJB)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

SP

ED146136

IMPROVING INSTRUCTION THROUGH STUDENT OBSERVATION OF TEACHING METHODS

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Pennsylvania Department of Education
1977

SP011 776

IMPROVING INSTRUCTION THROUGH STUDENT OBSERVATION OF TEACHING METHODS

James R. Masters
Pennsylvania Department of Education
September 1977

Commonwealth of Pennsylvania
Milton J. Shapp, Governor

Department of Education
Caryl Kline, Secretary

Bureau of Information Systems
Seon H. Cho, Director

Division of Research
Robert B. Hayes, Director

Pennsylvania Department of Education
Box 911
Harrisburg, PA 17126

ACKNOWLEDGMENTS

The author wishes to express appreciation to the administration of the Fairfield, Newport, West Shore and East Pennsboro school districts and the Cumberland-Perry Area Vocational Technical School for permitting this study to take place in their schools.

The conscientious efforts of the teachers and students who participated are also much appreciated.

Finally, this study would never have been performed were it not for the desire of the Student Advisory Board to produce an observation of teachers instrument of the highest possible quality. In particular, those students who served on the Tenure/Teacher Evaluation Committee are to be commended for their work in developing and refining the Student Observation of Teachers and Teaching Techniques (StOTT). The students listed below were instrumental in the development of the StOTT.

Victoria Beers	Franklin Area High School
Lee Byron	State College Area High School
Bryan Carlin	Berwick Area High School
Edie Cohen	Aliquippa High School
LuAnn Inman	Troy Senior High School
Judy Iskovitz	Taylor Allderdice High School
Judy Nolley	Pottsville Area High School
Tom Pearce	Grove City Area High School
Pam Ray	Redbank Valley High School
Pat Steigler	Palmerton Area Senior High School
Charles Wilson	North Penn High School

Table of Contents

	Page
Acknowledgments	iii
List of Tables	v
SUMMARY	vi
INTRODUCTION	1
THE STUDENT OBSERVATION OF TEACHERS AND TEACHING TECHNIQUES (StOTT) INSTRUMENT	1
Development of the Instrument	1
Description of the Instrument	2
DESIGN OF THE STUDY	3
RESULTS OF THE PRESENT STUDY AND OF PAST STUDIES	4
Examination of the StOTT Subscales	4
USE OF THE StOTT BY HIGH SCHOOL TEACHERS	21
REFERENCES	26
APPENDIX	29

List of Tables

		Page
1.	Reliabilities and Mean Item Ratings for Each StOTT Subscale	4
2.	Correlations Among the StOTT Subscales	5
3.	Stability Coefficients for the StOTT	8
4.	Comparisons of Ratings Given Two Teachers by the Same Class	12
5.	Comparisons of Ratings Given the Same Teacher by Two Classes	14
6.	Correlations between StOTT Ratings and Both Student and Teacher Characteristics	16
7.	Correlations between Mean StOTT Ratings Given a Teacher and Both Student and Teacher Characteristics	16
8.	Correlations between StOTT Ratings and Interest in School and Learning Scores	17
9.	Mean Item Ratings Obtained on Each Subscale by All 36 Teachers in the Study	19
10.	Percentages of Teachers Whose Mean Item Scores Fell at Different Levels on Subscales	24
11.	Percentages of Teachers Whose Mean Item Scores Fell at Different Levels on Factors	25

SUMMARY

Can students help teachers improve their instructional methods? This question was asked in the study described in this document.

In the study 925 high school students were asked to use the Student Observation of Teachers and Teaching Techniques (StOTT) instrument to give their opinions about the strengths and weaknesses of their teachers' methods. The StOTT instrument (see pages 28-30) was developed by high school students who served on the Student Advisory Board to the Secretary of Education of Pennsylvania.

The results of the study indicated that the StOTT instrument is a valuable one for use in individual teachers' classrooms. Student opinions about the same teacher's methods were found to vary little when measured on two different occasions a month apart. Also, students seemed capable of using the StOTT to identify what they believed to be the specific strengths and weaknesses of individual teachers.

The results indicated, however, that the StOTT should not be used to evaluate and compare teachers. This is because student opinions about individual teachers seemed to be influenced by students' own general feelings about school and teachers and their own abilities to succeed in school.

A summary of the results of other studies dealing with student observation of teaching methods is also included. The StOTT results generally agree with the results obtained in these studies.

INTRODUCTION

Most high school teachers have probably, at one time or another, made some attempt to gather student opinions about their teaching methods. Teachers who have used student feedback instruments would generally agree that information collected in this way is valuable. Hayes, Keim and Neiman (1967), for example, reported that over 90 per cent of the high school teachers in their study believed useful information was obtained from the student feedback instrument they were asked to use.

But how accurately can students assess teaching methods? Should a teacher make changes in his or her methods on the basis of what is found out from a student feedback instrument? Some educators argue that high ratings of a teacher by students simply indicate the teacher is popular. Others contend that students are highly changeable in their moods and opinions, and that ratings given during one week could differ greatly from those given during another. A third argument against the use of student feedback instruments is that the subject taught by a teacher has a great deal to do with his or her ratings. If the subject is an unpopular one, the teacher's ratings will suffer.

This study attempted to answer some of the questions posed above. The major purpose of the study was to examine the reliability and validity of an instrument developed to provide student feedback to high school teachers about their methods. In addition, the results of past investigations dealing with student feedback instruments were compiled to make a general statement about the strengths and weaknesses of such instruments.

THE STUDENT OBSERVATION OF TEACHERS AND TEACHING TECHNIQUES (StOTT) INSTRUMENT

Development of the Instrument

The instrument used in the study was designed to measure high school students' perceptions of their teachers. This questionnaire, called the Student Observation of Teachers and Teaching Techniques (StOTT) instrument, was developed by high school students who served on the Student Advisory Board to the Pennsylvania Department of Education. Their purpose for developing such an instrument was to improve teaching in Pennsylvania schools. This instrument, if proven effective, would be made available to high school teachers who wished to obtain student feedback on the strengths and weaknesses of their methods.

Development of the instrument began in October of 1973 at the first meeting of the Student Advisory Board. This board, composed of 11th and 12th grade student representatives from all parts of Pennsylvania, was formed to advise the Secretary of Education of student opinions about Pennsylvania's educational system. A Tenure/Teacher Evaluation Committee was set up by this board to study how students might help improve teaching. This committee decided to find or develop a feedback instrument of proven worth. After examining a number of questionnaires used in the United States, the committee, assisted by Department of Education advisers, began work on the development of their own instrument.[1]

---

[1] Waldo G. Weaver of the Division of Program Planning and Development served as the major adviser of this committee. Suggested procedures for using the StOTT and materials for preparing a StOTT profile chart can be obtained from him.

By the end of the 1973-74 school year, questions for an instrument had been written. These were taken into high school classrooms in nine school districts where students were asked to examine the clarity and appropriateness of each question. Teachers and administrators also were asked to examine the instrument at this time. Revisions in wording and the deletion of one question resulted from this tryout.

During the 1974-75 school year, the revised instrument was taken into classrooms again. This time students both reacted to the appropriateness and clarity of the questions and indicated their observations about their teachers. Again, the wording of a number of questions was changed. Some evidence of reliability and validity was found in this second tryout.

Since the instrument showed promise of being useful, the present study was designed to examine further its reliability and validity.

Description of the Instrument

The Student Observation of Teachers and Teaching Techniques instrument (shown in the Appendix) consists of 32 items. Each of the first 29 is included in one of five subscales. The final three items ask students to state whether the teacher they are rating has any distracting personal mannerisms, to give the reason they took the teacher's course and to tell whether they would recommend the course to another student.

The first 29 items are positive statements about a teacher's classroom behavior (e.g., "The teacher is willing to help you when you need it."). For each of these, a five-point scale is provided for students to mark their responses. The first, third and fifth points of the scale are defined for the student. These are, in the majority of cases, temporal (i.e., "Always," "Sometimes," "Never"), but for certain items a somewhat different scale is employed, for example:

The material on tests was covered by the teacher prior to the test.

5	4	3	2	1
Thoroughly		Moderately		Not At All

When examining the results of the class's rating, a teacher is encouraged to look at both the results for individual items and the means obtained for each of the five subscales. These subscales are: Teacher-Student Relations (items 1-6), Grades and Testing (items 7-12), Materials (items 13-15), Teacher Personality (items 16-18) and Teaching Methods and Techniques (items 19-29).

DESIGN OF THE STUDY

This study was carried out during the 1975-76 school year. Included were an examination of the stability of student ratings of the same teacher over time and an investigation of the extent to which student ratings of a teacher were influenced by extraneous factors, such as reasons for taking the teacher's course and general attitudes toward school and teachers.

To obtain as much information as possible from the tryout without using a great deal of any teacher's instructional time, three distinct groups were used. In all, 36 teachers participated, some of whom were rated by more than one class. A total of 925 10th, 11th and 12th grade students from 45 classes and five schools took part in the study. Two schools were suburban, two were rural and one was a vocational-technical school.

In the study, Group 1 consisted of 18 teachers whose classes filled out the StOTT, waited a month, then filled out the instrument again. These students also answered a questionnaire asking additional information about themselves, such as their sex, curriculum, reason for taking the course and average grade (both for the teacher's course and for their major courses). This group was used to examine both the stability of student ratings of the same teacher over time and the influence of extraneous factors on student ratings of a teacher.

Group 2 consisted of eight teachers, each of whom asked two of their classes to rate them. Both classes were of the same grade level and were taking the same subject. In this way, information was gathered on the extent to which ratings of a teacher were a function of the particular class chosen for the rating task. In addition, as a validation study of the instrument, class 1 responded to a forced-choice version of certain items in the StOTT. Class 2 responded to the instrument used in Pennsylvania's statewide assessment program (Educational Quality Assessment) to measure 11th grade students' interest in school and learning. The latter instrument was used to determine how much influence students' general attitudes toward school and teachers had on their ratings of an individual teacher.

Group 3 consisted of nine classes, each of which rated two of their teachers. Comparisons of the mean ratings given two teachers by the same class were carried out to determine if, in fact, students were capable of using the instrument to pick out specific strengths and weaknesses of their teachers' methods. If classes were found to give similar ratings to each of their teachers, the worth of the StOTT instrument would have to be seriously questioned.

RESULTS OF THE PRESENT STUDY AND OF PAST STUDIES

Examination of the StOTT Subscales

Before many of the major questions of the study could be answered, the data gathered on the StOTT subscales were carefully examined. If the five subscales were unreliable, there would be little worth in scoring the StOTT in this way. Also, if the correlations among the subscales were very high, this would seem to show that, rather than measuring five basic areas, the instrument was measuring only one general area--a kind of global indication of student opinions of their teachers. If this were the case, the worth of the StOTT in examining specific teaching methods would have to be seriously questioned.

Data from all 925 high school students who filled out the StOTT were used to examine the subscales. Table 1 shows the internal consistency reliabilities (coefficient alphas) computed for each of the subscales. This type of reliability estimate examines the degree to which the items in a subscale or test correlate with each other. The assumption in forming a subscale is that the items which make it up have much in common. By adding up the scores of these items, a subscale score which has meaning as a general assessment in an area can be obtained. As shown in Table 1, reliabilties of the five subscales ranged from .51 (Materials) to .80 (Teacher-Student Relations). In addition, each item of the StOTT was found to correlate more highly with its own subscale than with any other subscale. These results support the use of the subscales proposed. The reliability of the Materials subscale is somewhat low, however. Results on this subscale should be interpreted cautiously, since a general measure in this area would seem to require the use of more items than the three used.

Also shown in Table 1 are the mean item scores given in each subscale area by the 925 students. (A scale ranging from 1 to 5 was used in scoring each item.) For the total group of students, the highest ratings were for the items in the Grades and Testing subscale and the items in the Materials subscale. The lowest ratings were for Teaching Methods and Techniques.

Table 1

Reliabilities and Mean Item Ratings for Each StOTT Subscale

Subscale	No. of Items	Mean Item Rating	Coefficient Alpha
Teacher/Student Relations	6	4.12	.80
Grades and Testing	6	4.27	.65
Materials	3	4.28	.51
Teacher Personality	3	4.13	.71
Teaching Methods and Techniques	11	3.68	.79

Table 2 shows the correlations obtained among the five subscale areas. The highest correlation, .70, was between the Teacher Personality and Teacher-Student Relations subscales. The results in Table 2, then, indicate that each subscale measures something somewhat unique. Students responding to the StOTT did not appear to be responding in a general way, but rather seemed capable of rating the same teacher high in one area and low in another.

Table 2

Correlations Among the StOTT Subscales

	Teacher-Student Relations	Grades/ Testing	Materials	Personality
Grades/Testing	.54			
Materials	.38	.42		
Personality	.70	.50	.30	
Methods	.62	.63	.47	.64

The examination of the reliability and uniqueness of the StOTT subscales produced positive results. It was, therefore, possible to use these subscales in addressing additional questions about the instrument. Because similar questions have been addressed in past investigations of other student feedback instruments, an attempt was made to summarize these results along with those of the present study.

Below are questions which have been asked in investigations dealing with student feedback instruments. Following each question is a summary of results obtained both in the present study and in past studies. Although much of the work in this area has taken place at the college level, it was felt that, since the StOTT was developed for use in high schools, it would be most meaningful to limit the discussion as much as possible to educational levels below the college level.

1) What do students expect of their teachers? What do they feel are characteristics of good teachers?

The present study did not address this question, but it has been addressed by other investigators. Probably the most extensive research effort in this area was that of Kenny, Hentschel and Elpers (1972). In this study, groups of 70 to 150 students at four levels of education (grades 3, 7, 10 and college) were asked to list three qualities of good teachers and three characteristics of poor teachers. Student responses were sorted into one of 18 categories, and the percentages of responses which fell into each category were determined.

At all four grade levels, the category found to be most important was attitude toward students; it was mentioned either most often or second most often at each level. The types of responses categorized into this area, however, differed somewhat as a function of student grade level. Elementary school students felt that good teachers were patient and understanding and did not yell a great deal. Middle school students wanted teachers to be helpful to them when they needed it and to try to understand them. They disliked teachers who yelled and were critical. High school students felt that the best teachers cared for students and did not unjustly criticize them or make them feel inferior. At the college level, students believed good instructors to be those who gave individual attention, remembered names and gave breaks when needed.

At each level, students also mentioned factors about teachers' methods which were important to them. At the elementary level, students best liked teachers who provided creative free time and taught in a variety of ways. Middle school students felt that good teachers gave meaningful assignments, used a variety of teaching techniques and were capable of giving good explanations. Poor teachers were believed to be unfair to students and to give assignments which were not worthwhile. High school students listed a number of characteristics of good teachers. Good teachers were said to use a variety of teaching methods, to have the ability to stimulate interest, to provide clear explanations, to have a sense of humor and to have an interest in the subject being taught. Poor teachers were characterized as being unable to control a class properly, either by providing too little discipline or discipline which was too rigid. Other characteristics of poor teachers were unfairness in dealing with students and inability to provide clear explanations. College students felt that good instructors had the ability to stimulate interest, a realistic attitude toward their subject, a sense of humor and a comprehensive knowledge of their subject. Poor college teachers were said to speak over the heads of students, to have little interest in what they were teaching and to treat students unfairly.

In a similar study, Tofelson (1973) asked 1,643 high school students to list characteristics they believed made a teacher effective in the classroom. The characteristics this group of students mentioned were similar to those listed by high school students in the study described above. They were tolerance, flexibility, respect for students, skill in presenting subject matter and enthusiasm. Rural, urban and suburban students did not differ in their responses.

Thus, high school students do seem to feel that many of the areas included in the StOTT have much to do with the effectiveness of teachers. A teacher's attitude toward them seems very important to them. Also important, however, are a teacher's abilities to present material and to control a class.

2) Are student ratings of a teacher stable over time? Do they fluctuate to such an extent that a teacher can have little confidence in their reliability?

If a teacher is to take seriously criticisms obtained through a student feedback instrument, he or she must believe that students are capable of giving their general impressions rather than those which they happen to have on a particular day. Therefore, the question above was examined closely in the StOTT tryout.

To answer the question, 18 classes were asked to fill out the StOTT for a teacher, to wait a month and then to fill it out again for the same teacher. Correlations between the ratings given the same teacher on the two occasions provided an estimate of stability.

Table 3 shows rather conclusively that student ratings of the same teacher had a great deal of stability. For the total group of 334 students, correlations were above .70 for four of the subscales. Again, the Materials subscale had the least positive results, but it is possible that actual changes in materials used were as much responsible for this as instability in student ratings. Mean ratings of teachers on the subscale were extremely stable. Teachers who received the highest mean scores for a subscale on the first administration of the StOTT also tended to receive the highest mean scores on the second administration.

It was believed that students experiencing learning difficulties in school might, because of low motivation and inability to read well, produce ratings which had a degree of instability. Therefore, a comparison was made of the stability of ratings given by students averaging grades of A or B in their major courses with that of students averaging grades of C and below. As seen in Table 3, the results of this comparison supported the findings cited above. Students earning grades of C and below did not appear to give ratings which were any less stable than were those of students earning grades of A or B.

Table 3

Stability Coefficients for the StOTT

Group	No. of Students	Teach.-Stud. Relations	Grades/ Testing	Materials	Person-ality	Methods
Students Averaging Grades of C or Below	109	.80	.67	.60	.78	.70
Students Averaging Grades of A or B	208	.79	.79	.61	.72	.73
ALL STUDENTS	334	.79	.76	.60	.75	.71
Mean Ratings Given to 18 Teachers		.93	.92	.79	.91	.76

A number of other investigators have also examined the stability of student ratings of a teacher. Davidoff (1970) found that ratings of high school teachers were stable over a two-day period. DeAmico (1973) reported that, for specific items in a 16-item rating scale, the average rating for the teachers in his study was highly similar over a two-year period. At the elementary level, Hayes, Keim and Neiman (1967) computed coefficients of consistency ranging from .58 (with 21 weeks between ratings) to .85 (with five weeks between).

Both the present study and past studies, then, indicate that students are capable of using feedback instrument to provide ratings which have a great deal of stability. Teachers can have some confidence that students are using these instruments to give their general ratings of a teacher's methods rather than ratings which merely indicate their feelings on a particular day.

3) Is there any validity to student ratings of a teacher's methods? Do students approach the rating task in a thoughtful, careful way? Do student ratings of a teacher's effectiveness agree with other measures of the same teacher's effectiveness?

The present study addressed this question primarily from the standpoint of determining the extent to which students made thoughtful, reasoned decisions about their teachers' methods. Eight classes of students were asked to fill out the StOTT for a teacher. Then, without looking at these ratings, they were asked to respond to certain of the questions from the StOTT again, this time in a forced-choice format. Agreement between ratings given in the two different ways would provide evidence that students were making thoughtful decisions when responding to the StOTT.

The forced-choice instrument was made up of four items from the Teacher-Student Relations subscale and four items from the Methods subscale. Each Teacher-Student Relations item was paired with each of the Methods items, producing a total of 16 pairs. Students responding to the instrument were asked which statement of each pair was more true of the teacher they were rating. A student's score on the instrument was the number of times he or she had chosen a Teacher-Student Relations item over a Methods item.

In addition, for the same 16 pairs, the same student's item scores on the StOTT were examined. The number of times he or she had given a higher rating to a Teacher-Student Relations item than to a Methods item was computed (e.g., a rating of 5 on item 1 and a rating of 4 on item 23 would be counted in a similar way to choosing item 1 over item 23 in a forced-choice pairing).

Correlations between scores obtained in the two ways indicated a great deal of consistency. For the total group of 171 students, a correlation of .70 was computed. For six of the eight classes, correlations of .59 or greater were found.

Students, therefore, did appear to be responding to each item of the StOTT in a thoughtful way. The difficult question of whether student perceptions of a teacher's strengths and weaknesses are accurate was not addressed by the present study, since no criterion of effectiveness was readily available. However, students at least seemed capable of taking the task seriously and of making discriminations among the five points of the StOTT item scales which had meaning to them.

Other investigators have examined the validity of student observations of teachers in their studies. Hayes, Keim and Neiman (1966, 1967) addressed this issue at both the elementary level and the high school level. In 1966 they found a correlation of .35 between principals' ratings and students' ratings of the same high school teachers. At the elementary level, a correlation of .27 was computed between principals' ratings and students' ratings of sixth grade teachers. The almost unanimous opinion of these teachers was that their pupils had accurately rated their teaching.

Veldman and Peck (1967) uncovered similar results in a study in which junior and senior high school students rated student teachers. Ratings given these teachers by their university supervisors agreed well with student ratings of their teaching.

A number of researchers have attempted to determine whether teachers rated most highly by their students are actually those from whom students learn the most. Davidoff (1970) cited Fortune (1965), Morsh, Burgess and Smith (1956) and Podlogar, Rosenshine and Gage (1967) as investigators who found evidence that students could pick out and give high ratings to teachers who caused students to learn the most. However, in his own study, Davidoff found no consistent relationship between student ratings of teacher behavior and the same students' gains in achievement. Bryan (1963) reported similar results for high school chemistry classes. Although he uncovered no significant relationship between students' achievement gains and their ratings of teachers, he did find a high correlation between teacher prestige with students and the development of interest in and a liking for chemistry. The image students had of these chemistry teachers was similar to that held by administrators and parents, and was stable from year to year.

Students, then, do seem capable of making reasoned judgments about teacher behavior. These judgments, in many cases, seem similar to those of others who have contact with the same teachers. Nevertheless, research dealing with the relationship between student ratings of teachers and student achievement has produced no clear pattern. Teachers rated highest were found to produce the greatest student learning gains in some studies but not in others.

4) Are feedback instruments capable of measuring ways in which teachers differ? If the same class is asked to rate more than one teacher, will these ratings actually differ?

This question was believed a key one in determining the usefulness of the StOTT. In any rating task, a portion of the score must be attributed to the unique perceptions of those doing the rating. If this portion becomes too large, differences in ratings given must be viewed primarily as a function of those doing the rating rather than a function of the person being rated. If this occurred for the StOTT, it would be expected, for example, that the same class would view all teachers similarly and would give them similar ratings.

In the study, nine classes were asked to fill out the StOTT for two of their teachers. Classes chosen for this task were those in which all or almost all of the class members were enrolled in classes taught by the same two teachers (e.g., all students had the same English teacher and the same history teacher).

Correlated t tests were used to compare the scores of the two teachers rated by the same class. Table 4 shows the mean item scores obtained for each subscale. Also shown are the t values computed.

It can be seen in Table 4 that the StOTT was capable of measuring ways in which teachers differed. Every class rated one of their two teachers higher than the other on at least one subscale. Of special interest are the results for the first two classes shown. These classes rated one of their teachers higher on the Grades and Testing subscale and the other teacher higher on the Personality subscale. This is an indication that these classes were not providing some global rating of their teachers but were reacting to the specific content of each item.

5) Do classes differ in their ratings of the same teacher? If a teacher would like to use a student feedback instrument, is it necessary for him or her to use it with more than one class?

Teachers teaching at the secondary level generally teach a number of different sections of students, some of which may be enrolled in exactly the same course. From a practical standpoint, it would seem to make sense to use a feedback instrument with only one of these classes and to assume that ratings of the other sections would be similar. If, on the other hand, it were possible for the ratings of a teacher to vary a great deal as a function of the particular class chosen for the rating task, misleading results could be gotten from the use of only one class's ratings.

To investigate whether classes could differ in their ratings of the same teacher, eight teachers asked two of their classes to fill out the StOTT. They were instructed to choose for this task sections enrolled in exactly the same subject at the same grade level.

Table 5 shows the mean item scores obtained on each subscale for each of the classes which rated the same teacher. Also shown are F values computed for the analyses of variance performed to compare these means.

As shown in Table 5, for three of the eight teachers, classes did differ in their ratings. In all three cases, one of the classes seemed to view the teacher more positively, regardless of the subscale. Thus, for almost one-half of the teachers, the results of StOTT ratings did differ as a function of the particular class asked to fill out the form, even though each class was enrolled in exactly the same course.

Table 4

Comparisons of Ratings Given Two Teachers by the Same Class

Class	No. of Stud.	Teacher-Student Relations			Grades and Testing			Materials			Personality			Methods		
		Mean Item Rating			Mean Item Rating			Mean Item Rating			Mean Item Rating			Mean Item Rating		
		Teach. A	Teach. B	t Value	Teach. A	Teach. B	t Value	Teach. A	Teach. B	t Value	Teach. A	Teach. B	t Value	Teach. A	Teach. B	t Value
1	13	4.14	3.96	0.76	4.13	4.39	-2.46*	4.23	4.18	0.49	4.26	3.72	3.41**	4.01	3.92	0.78
2	22	3.71	3.74	-0.19	4.27	3.83	2.87**	4.17	3.64	2.98**	3.42	4.26	-6.27**	3.86	3.12	4.98**
3	24	4.26	3.86	3.65**	4.17	4.03	0.94	4.49	4.00	3.61**	3.97	3.74	1.87	3.92	3.78	1.34
4	14	4.07	4.43	-1.84	4.22	4.48	-2.15	3.98	4.36	-2.58*	3.74	4.67	-5.64**	3.68	4.12	-4.89**
5	16	4.71	4.38	3.65**	4.50	3.61	6.63**	4.60	4.21	2.64*	4.50	4.50	0.00	4.34	3.92	2.49*
6	18	4.64	4.42	1.70	4.44	3.91	2.50*	4.46	4.41	0.41	4.54	4.57	-0.21	4.09	4.00	0.51
7	25	4.14	4.30	-1.28	3.99	4.23	-2.17*	4.00	4.29	-2.40*	4.20	4.36	-1.33	3.95	4.07	-1.42
8	17	3.74	4.00	-2.62*	3.57	4.10	-3.39**	3.67	4.04	-2.12*	3.78	4.04	-1.28	3.50	3.87	-3.24**
9	15	4.29	3.92	2.71*	4.50	3.72	3.69**	4.22	4.13	0.53	4.44	4.31	1.25	4.09	3.87	1.38

*Significant at beyond the .05 level.

**Significant at beyond the .01 level.

Only one other study below the college level was found to have addressed a question similar to the one stated above. Veldman and Peck's (1967) overall conclusion, however, differed from the findings for the StOTT. In their study, classes which rated the same student teacher gave similar ratings. A total of 50 student teachers were rated by more than one class. At the college level, VanNelson and Cossart (1975) uncovered results which also indicated that classes saw the same teacher in a similar way. Their investigation was carried out in the classes of mathematics and science professors.

The fact that the results for the StOTT differed from those of past studies could be explained in a number of ways. For example, it is possible that the StOTT is capable of measuring finer differences than were measured by instruments used in past studies. Or it may be that if more than eight classes were used in the study, the general trend would be toward the finding of similar ratings across classes. In any case, this would seem a matter of judgment for teachers using the StOTT. If classes being taught by a teacher differ in some way or have different rapport with the teacher, the present study has demonstrated that it is possible for these classes to rate the teacher differently on any of the subscales of the instrument.

6) **Are there factors other than a teacher's behavior which influence** the type of rating given him or her by students? Can a teacher receive either high or low ratings simply because of the type of course being taught or the type of student enrolled in his or her class?

This question has received a great deal of attention in past studies of student feedback instruments. Much of this work has been carried out at the college level, where, in some cases, attempts have been made to compare the effectiveness of instructors by asking students to fill out observation forms for most or all of their instructors.

The present study addressed this question primarily from the standpoint of attempting to help teachers to understand the extent to which factors other than their own behavior could be responsible for the ratings they receive. The StOTT was not developed as an instrument for comparing the effectiveness of teachers; rather, it was developed as a tool for teachers to use in their own classrooms.

Table 5

Comparisons of Ratings Given the Same Teacher by Two Classes

Teacher	Number of Students		Teacher-Student Relations			Grades and Testing			Materials			Personality			Methods		
			Mean Item Rating			Mean Item Rating			Mean Item Rating			Mean Item Rating			Mean Item Rating		
	Class 1	Class 2	Class 1	Class 2	F Value	Class 1	Class 2	F Value	Class 1	Class 2	F Value	Class 1	Class 2	F Value	Class 1	Class 2	F Value
A	26	26	3.57	3.87	3.91	3.95	4.20	4.82*	3.38	4.04	12.89**	3.49	4.06	7.59**	3.43	3.76	6.00*
B	19	24	3.61	3.67	0.05	3.92	3.49	2.96	4.04	4.00	0.02	3.95	3.72	0.78	3.68	3.53	0.52
C	28	21	3.88	3.70	0.92	3.97	3.96	0.00	4.38	4.30	0.30	3.80	3.60	0.91	3.58	3.71	0.47
D	21	15	4.14	3.63	4.11	4.41	3.91	9.56**	4.35	4.18	0.80	4.22	3.76	6.23*	4.13	3.58	8.28**
E	23	14	3.48	3.31	0.72	4.39	4.44	0.10	4.48	4.48	0.00	3.48	3.45	0.01	3.90	3.67	1.32
F	9	20	4.02	3.90	0.36	4.41	4.19	1.24	4.22	3.92	2.14	4.00	4.03	0.01	3.60	3.81	0.64
G	26	27	4.20	4.66	13.15**	4.34	4.72	11.61**	4.06	4.38	3.73	4.58	4.89	12.56**	4.18	4.57	15.18**
H	19	15	4.49	4.60	0.44	4.63	4.75	1.20	4.32	4.38	0.12	4.49	4.62	0.50	4.04	4.08	0.09

*Significant at beyond the .05 level.

**Significant at beyond the .01 level.

In the StOTT tryout, 18 classes were asked to fill out both a form giving information about themselves and the StOTT. They were to indicate their sex, their grade level, their curriculum, their reason for taking the course taught by the teacher they were rating, their average grade in that course and their average grade for all of their courses. Also, eight classes responded to both the StOTT and the instrument used in Pennsylvania's statewide assessment program (Educational Quality Assessment) to measure 11th grade students' interest in school and learning. This instrument contains subscales measuring attitudes toward learning, toward school and toward teachers. [See Hertzog (1976) for more information about the instrument.]

Tables 6 and 7 summarize the relationships found between the student and teacher characteristics measured and student ratings of teachers. In Table 6, many significant relationships are shown for the total group of 334 students. However, all of these would be described by Cohen (1969) as effects of small or medium size, i.e., effects which met the criterion of statistical significance but which were of small magnitude. For example, if an attempt were made to predict teacher ratings on the Personality subscale on the basis of the highest relationship shown, .29 for Teacher Sex, only $(.29)^2 = 8.41$ per cent of the variance in these ratings could be predicted. Over 90 per cent of these ratings would be unexplained by the fact that the teacher was either male or female.

To construct Table 7, means were computed on each student characteristic variable for each of the 18 classes. These means, along with the Teacher Sex variable, were correlated with the mean ratings of each teacher on the five StOTT subscales.

Many of the correlations shown in Table 7 are significant and of rather large magnitude. Thus, much of the reason for the finding of significant relationships for the total group appears to be that there was a tendency for classes which gave teachers high ratings to differ from classes which gave low ratings. Classes which gave high ratings tended to be made up of a higher proportion of female students and to include students who were earning higher grades both in the rated teacher's course and in other courses. (The correlation between these two variables was .50.) Female teachers tended also to be rated higher than male teachers. Ratings which were significantly related to Teacher Sex and to student characteristics were limited almost exclusively to the Teacher-Student Relations subscale and the Personality subscale.

Table 8 contains correlations computed between StOTT student ratings and their scores on the Interest in School and Learning instrument. These two instruments were administered to 161 students from eight classes.

Table 6

Correlations Between StOTT Ratings and Both Student and Teacher Characteristics (N = 334)[1]

Variable	Teacher-Student Relations	Grades and Testing	Materials	Personality	Methods
Student Grade Level (10,11,12)	.04	-.03	.03	.15**	.08
Student Sex (1 = Male, 2 = Female)	.18**	-.01	.11*	.19**	.03
Student Grade in Teacher's Course (A = 5, B = 4, etc.)	.27**	.05	.13**	.23**	.19**
Student Average Grade for Major Courses (A = 5, B = 4, etc.)	.19**	-.02	.13**	.16**	.03
Difference Between Student Grade in Course and Grade in Major Courses[2]	.16**	.08	.03	.14**	.20**
Student Reason for Taking Course (2 = Wanted To, 1 = Did Not Want To)	.09*	.11**	.07	.07	.17**
Teacher Sex (1 = Male, 2 = Female)	.23**	.07	-.02	.29**	.07

Table 7

Correlations Between Mean StOTT Ratings Given a Teacher and Both Student and Teacher Characteristics (N = 18)[1]

	Teacher-Student Relations	Grades and Testing	Materials	Personality	Methods
Student Grade Level	.20	-.06	-.06	.34	.21
Student Sex	.68**	-.07	.41*	.57*	.12
Student Grade in Teacher's Course	.63**	-.01	.31	.54**	.19
Student Average Grade for Major Courses	.60**	-.03	.21	.57**	.04
Difference Between Student Grade in Course and Grade in Major Courses[2]	.14	.02	.39	.04	.21
Student Reason for Taking Course	.03	.43*	.08	-.06	.35
Teacher Sex	.41*	.00	.09	.51*	.21

*Significant at beyond the .05 level.
**Significant at beyond the .01 level.

[1]Since hypotheses about the direction of differences were not stated for Student Grade Level, Student Sex and Teacher Sex, two-tailed tests were carried out for these. For all other variables one-tailed tests were performed.

[2]Positive values for this variable indicated that the teacher being rated gave students grades which were higher than those being received in their major subject courses.

Almost all of the correlations shown in Table 8 are significant. Students who held the most positive attitudes toward learning, toward school and toward teachers did tend to give teachers the highest StOTT ratings.

Table 8

Correlations between StOTT Ratings and Interest in School and Learning Scores (N = 161)

Interest in School and Learning Subscale	Teach.-Stud. Relations	Grades/ Testing	Materials	Personality	Methods
Learning	.18**	.08	.30**	.04	.28**
School	.24**	.18**	.29**	.16*	.30**
Teachers	.35**	.33**	.34**	.23**	.40**
Total Instrument	.30**	.22**	.36**	.18**	.38**

*Significant at beyond the .05 level.
**Significant at beyond the .01 level.

Table 9 was constructed to examine the relationship between student ratings of a teacher and the subject taught by the teacher. The number of teachers representing each subject area was felt to be too small for a statistical test to be performed. However, informal examination of Table 9 seems to indicate that the subject taught did not have a major impact on ratings. For subject areas in which a number of teachers were rated, ratings varied rather widely as a function of individual teachers.

Past research dealing with relationships between student characteristics and the same students' ratings of their teachers has produced mixed results. On the one hand, Thompson (1974) found no significant relationship between ratings of high school teachers and their students' sex, grade level, grade-point average, expected course grade or degree of absenteeism. Shaw (1973) concluded that ratings of high school teachers were not influenced by the teacher's sex, academic degrees or years of teaching experience or by students' grade-point averages.

On the other hand, however, Smith and Brown (1976) uncovered results similar to those of the present study. They concluded that students' general attitudes toward teachers, their enjoyment of the subject matter and the grade they expected to receive all influenced ratings given a teacher. Jackson and Fuller (1966) reported that female students perceived teachers as more confident or posed. They also found that the socioeconomic status of students and teachers seemed to enter into student ratings.

Table 9

Mean Item Ratings Obtained on Each Subscale by
All 36 Teachers in the Study

Subject Area	Teacher	Teacher-Student Relations	Grades and Testing	Materials	Personality	Methods
Mathematics	A	4.39	4.46	4.32	4.46	3.81
	B	4.55	4.70	4.35	4.57	4.07
	C	4.26	4.17	4.49	3.97	3.92
Social Science	A	4.36	3.98	4.44	4.06	4.05
	B	4.60	4.55	4.44	4.79	4.41
	C	3.72	4.08	3.71	3.78	3.60
	D	3.79	3.96	4.34	3.70	3.65
	E	4.43	4.53	4.22	4.74	4.38
	F	3.94	3.78	3.84	3.99	3.73
	G	4.24	3.74	4.25	4.46	3.93
	H	3.90	4.22	4.08	3.66	3.81
	I	3.86	3.88	4.59	3.75	3.61
Foreign Language	A	4.59	4.82	4.73	4.60	4.43
	B	4.37	4.62	4.31	4.51	4.18
English	A	4.22	4.33	4.37	4.56	4.08
	B	4.30	4.48	4.23	4.24	4.14
	C	4.26	4.22	4.05	4.40	4.03
	D	4.36	4.31	4.36	4.41	4.15
	E	4.55	4.48	4.43	4.49	4.18
	F	4.14	4.13	4.23	4.26	4.01
	G	4.43	4.48	4.36	4.67	4.12
Science	A	4.55	4.08	4.01	4.59	4.07
	B	3.64	3.71	4.02	3.84	3.61
	C	3.73	3.83	3.64	4.26	3.12
Business	A	4.69	4.78	4.68	4.37	3.84
	B	4.59	4.09	4.56	4.74	4.14
	C	3.96	4.30	4.07	4.02	3.71
Vocational Education	A	4.03	4.25	4.22	3.89	4.08
	B	4.01	4.43	4.41	4.04	3.95
	C	3.60	4.54	4.60	3.42	4.11
	D	3.58	4.13	4.33	3.22	3.73
	E	3.88	4.51	4.57	3.43	3.95
	F	3.21	4.25	3.94	3.72	3.89
	G	3.85	4.38	4.38	3.96	3.93
	H	3.88	4.16	4.27	3.99	3.86
	I	3.40	4.42	4.48	3.47	3.79

Veldman and Peck (1967) found results which indicated that student ratings of teachers varied as a function of the subject area taught. The most highly rated teachers were those teaching physical education, home economics or business. Average ratings were received by teachers of social science or English. Fine arts teacher ratings were described as below average; those of foreign language teachers were quite low. Mathematics and science teachers received the lowest ratings.

Thus, some results of the present study agree with those of past studies; others do not. Student characteristics do seem to influence the types of ratings given teachers on the StOTT. It may be that instruments used in investigations in which this relationship was not found were different from the StOTT in some way.

7) Is there another useful way of scoring the StOTT? Can a statistical analysis of student responses to items suggest a valuable way of looking at the information gained from an administration of the StOTT?

The StOTT subscales were formed by placing items of like content on the same subscale. The present study provided empirical support for this structure.

It was felt, however, that there might be some value to investigating the structure more closely. To do this, a procedure called factor analysis was used. This statistical procedure examines the degree of correlation among variables or items. Its purpose is to reduce the number of variables one must deal with to some manageable number, such as three or four rather than 29. Factors are extracted and variables which are most highly related to each other are said to "load highly" or correlate with the same factor. Thus, for example, 10 of 29 items might load highly on the first factor extracted. This factor would be given a name dependent completely upon the content of the particular items which loaded highly on it. In the same way, other factors would be extracted, and those items which loaded highly on them would influence the name given these factors. By adding a person's scores on those items loading highly on the same factor, a total score on the factor can be obtained for the person.

Item data from all 925 students in the StOTT study were entered into a factor analysis.[1] Three meaningful factors resulted:

---

[1] Principal axis factor analysis was employed with squared multiple correlations as communality estimates (iterated until convergence was achieved). Six factors with eigenvalues greater than 1.0 were extracted, but the four-, five- and six-factor solutions were not readily interpretable, and each included a single-item factor. The scree test (Cattell, 1966) indicated that the three-factor solution was appropriate, and it was interpretable. A total of 40 per cent of the variance in the 29 items was accounted for by the three factors. Both oblique and varimax rotations were performed with highly similar results.

I. Student Perception of the Teacher (8 items)--made up of six Teacher-Student Relations items and two Personality items. Items included were numbers 1-6, 17 and 18.

II. Methods and Materials (16 items)--made up of all items included in Grades and Testing, all items in Materials and most Methods items. Items included were numbers 7-15, 22 and 24-29.

III. Teacher Ability to Interest Student (5 items)--made up of one Personality item (enthusiasm for the subject taught) and four Methods items. Items included were numbers 16, 19-21 and 23.

The first two factors extracted appear more general than the StOTT subscales. In essence, they combine similar StOTT subscales. The third factor, however, seemed to provide a new dimension. The items loading highly on it all pertain to a teacher's ability to interest students. Items loading highly on this factor measure student opinions about the meaningfulness of a teacher's lectures and discussions, the teacher's enthusiasm for his or her subject, the teacher's ability to stimulate discussion and the extent to which the teacher varies his or her methods of presentation. A score obtained on this factor would seem a valuable addition to those obtained on the five original StOTT subscales.[2]

USE OF THE StOTT BY HIGH SCHOOL TEACHERS

The Student Observation of Teachers and Teaching Techniques instrument appears highly useful for high school teachers. Student ratings of a teacher with the instrument were found stable over an interval of a month. Students seemed capable of making and repeating discriminations among the five scale points for each item. Classes were able to employ the StOTT to pick out ways in which they felt their teachers differed. The subscales of the instrument make sense conceptually and, in general, were found to be internally consistent.

The major problem uncovered in the study was that students of differing types tended to rate teachers differently. If the results of this part of the study are replicable (and, indeed, they are similar to those of Smith and Brown, 1976), teachers of students who like school and are doing well in it would be expected to routinely receive higher StOTT ratings than teachers of students who are "turned off" by school and are not doing well.

---

2 To examine the results of structuring the StOTT in terms of the factors extracted, each item was assigned to the factor for which its loading was highest and coefficient alpha reliabilities for the three factors were computed (using item scores from all 925 students). These were, for factors I, II and III, respectively, .86, .85 and .63. Correlations among the factors were: I with II, .66; I with III, .48; and II with III, .53.

These findings are not surprising. Students who hold generally poor attitudes toward teachers, for example, would not be expected to give high ratings to individual ones.

The results suggest a number of conclusions with respect to use of the StOTT by high school teachers. The following list of conclusions and suggestions was prepared on the basis of what was learned:

1) The StOTT should <u>not</u> be used to evaluate or compare high school teachers. The types of students being taught appear to influence ratings of teachers too greatly for this. The study demonstrated that it is even possible for two similar classes taking the same course from a teacher to disagree in their ratings.

2) The StOTT is a valuable instrument for teachers to use in their own classrooms. Teachers who are aware of the types of students they are teaching can take factors such as student attitudes into consideration when interpreting their ratings.

3) When interpreting the results of a StOTT administration, both results for individual items and for subscales appear valuable. In addition, the third factor extracted in the factor analysis provides a new dimension to StOTT scoring.

4) The StOTT can be used to measure changes in student opinions about a teacher's methods during the school year. The study examined the stability of ratings over a month's time, but the mean ratings received by a teacher were so stable that there would seem to be little loss in reliability when measuring changes over a longer period of time. In doing this, of course, a teacher would want to measure opinions when students are experiencing their usual day-to-day routines. For example, ratings given during the last few days of school might be influenced by students' elation that the school year is ending. It would seem better to measure these end-of-year feelings in late April or very early May.

5) The study found it was possible for two similar classes to disagree in their ratings of the same teacher. Thus, a teacher who uses the StOTT might want to administer it to more than one class. If low ratings on an item, a subscale or a factor are given by two or more classes, this would suggest that the teacher should look much more closely at this area than at an area for which low ratings were received from only one class.

Finally, as an aid to interpreting the results of StOTT administrations, Tables 10 and 11 were prepared. These tables show the percentages of the 36 teachers in the study whose mean item ratings on subscales and factors fell at differing levels.

These tables suggest that students tend to rate teachers higher in some areas than in others. As can be seen, for example, a mean item rating of 4.00 received on the Ability to Interest Students factor would be higher than that received by 72.23 per cent of the 36 teachers in the study. A mean item rating of 4.00 received on the Grades and Testing subscale, however, would be higher than that received by only 19.45 per cent of the 36 teachers. Therefore, if these 36 teachers are typical, teachers using the StOTT might expect to receive higher ratings on some subscales or factors than on others.

The results of the study, then, point to the conclusion that the StOTT instrument is capable of fulfilling the purpose for which it was constructed. Its value lies in providing feedback to individual teachers about their students' opinions of their methods. The instrument was not intended to be a device for comparing the effectiveness of teachers, and the results obtained indicate that it should not be used for this purpose.

Table 10

Percentages of Teachers Whose Mean Item Scores
Fell at Different Levels on Subscales
(No. of Teachers = 36)

Item Score Interval	Teacher-Student Relations			Grades and Testing			Materials			Teacher Personality			Teaching Methods and Techniques		
	No. of Teachers	%	% Below	No. of Teachers	%	% Below	No. of Teachers	%	% Below	No. of Teachers	%	% Below	No. of Teachers	%	% Below
4.90-5.00				0	0.00	100.00				0	0.00	100.00			
4.80-4.89				1	2.78	97.23	0	0.00	100.00	3	8.33	91.67			
4.70-4.79	0	0.00	100.00	2	5.56	91.67	1	2.78	97.23	2	5.56	86.11			
4.60-4.69	2	5.56	94.45	1	2.78	88.89	2	5.56	91.67	4	11.11	75.00			
4.50-4.59	5	13.89	80.56	4	11.11	77.78	3	8.33	83.34	5	13.89	61.11	0	0.00	100.00
4.40-4.49	2	5.56	75.00	6	16.67	61.11	6	16.67	66.67	1	2.78	58.33	2	5.56	94.44
4.30-4.39	5	13.89	61.11	4	11.11	50.00	9	25.00	41.67	3	8.33	50.00	1	2.78	91.66
4.20-4.29	4	11.11	50.00	4	11.11	38.89	6	16.67	25.00	0	0.00	50.00	0	0.00	91.66
4.10-4.19	1	2.78	47.22	4	11.11	27.78	0	0.00	25.00	3	8.33	41.67	7	19.44	72.22
4.00-4.09	2	5.56	41.66	3	8.33	19.45	5	13.89	11.11	4	11.11	30.56	7	19.44	52.78
3.90-3.99	3	8.33	33.33	2	5.56	13.09	1	2.78	8.34	2	5.56	25.00	5	13.89	38.89
3.80-3.89	4	11.11	22.22	2	5.56	8.33	1	2.78	5.56	4	11.11	13.89	5	13.89	25.00
3.70-3.79	3	8.33	13.89	3	8.33	0.00	1	2.78	2.78	1	2.78	11.11	4	11.11	13.89
3.60-3.69	2	5.56	8.33				1	2.78	0.00	0	0.00	11.11	4	11.11	2.78
3.50-3.59	1	2.78	5.56							3	8.33	2.78	0	0.00	2.78
3.40-3.49	1	2.78	2.78							0	0.00	2.78	0	0.00	2.78
3.30-3.39	0	0.00	2.78							1	2.78	0.00	0	0.00	2.78
3.20-3.29	1	2.78	0.00										0	0.00	2.78
3.10-3.19													1	2.78	0.00

Table 11

Percentages of Teachers Whose Mean Item Scores Fell at Different Levels on Factors (No. of Teachers = 36)

Item Score Interval	Student Perception of the Teacher			Methods and Materials			Teacher Ability to Interest Students		
	No. of Teachers	%	% Below	No. of Teachers	%	% Below	No. of Teachers	%	% Below
4.90-5.00									
4.80-4.89				0	0.00	100.00			
4.70-4.79	0	0.00	100.00	1	2.78	97.21			
4.60-4.69	3	8.33	91.68	0	0.00	97.21			
4.50-4.59	4	11.11	80.57	3	8.33	88.88			
4.40-4.49	3	8.33	72.24	3	8.33	80.55	0	0.00	100.00
4.30-4.39	5	13.89	58.35	7	19.44	61.11	1	2.78	91.67
4.20-4.29	2	5.56	52.79	3	8.33	52.78	3	8.33	88.89
4.10-4.19	2	5.56	47.23	7	19.44	33.34	3	8.33	80.56
4.00-4.09	0	0.00	47.23	3	8.33	25.01	3	8.33	72.23
3.90-3.99	4	11.11	36.12	4	11.11	13.90	7	19.44	52.79
3.80-3.89	4	11.11	25.01	2	5.56	8.34	4	11.11	41.68
3.70-3.79	4	11.11	13.90	2	5.56	2.78	6	16.67	25.01
3.60-3.69	1	2.78	11.12	0	0.00	2.78	5	13.89	11.12
3.50-3.59	0	0.00	11.12	1	2.78	0.00	1	2.78	8.34
3.40-3.49	2	5.56	5.56				1	2.78	5.56
3.30-3.39	0	0.00	5.56				0	0.00	5.56
3.20-3.29	1	2.78	2.78				1	2.78	2.78
3.10-3.19	1	2.78	0.00				1	2.78	0.00

REFERENCES

Bryan, R. B. Reactions to Teachers by Students, Parents and Administrators. Kalamazoo, Michigan: Western Michigan University, 1963.

Cattell, R. B. The scree test for the number of factors. Multivariate Behavioral Research, 1(2):245, 1966.

Cohen, J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1969.

Davidoff, S. H. The development of an instrument designed to secure student assessment of teaching behaviors that correlate with objective measures of student achievement. Paper presented at the American Educational Research Association Annual Meeting, Minneapolis, 1970.

DeAmico, G. Second Year Report: A Student's Rating Scale of a Teacher. Sheboygan Public Schools, Wisconsin, 1973.

Fortune, J. C. Generalities of the Ability to Explain. Unpublished doctoral dissertation. Palo Alto, California: Stanford University, 1965.

Hayes, R. B., Keim, F. N. and Neiman, A. M. The Effects of Student Reactions to Teaching Methods. USOE-Project No. 5-0096, 1966.

Hayes, R. B., Keim, F. N. and Neiman, A. M. The Effects of Student Reactions to Teaching Methods. USOE-Project No. 6-2056, 1967.

Hertzog, J. F. Getting Inside the EQA Inventory. Harrisburg: Pennsylvania Department of Education, 1976.

Jackson, M. L. and Fuller, F. F. Influence of social class on students' evaluations of their teachers. Proceedings of the 74th Annual Convention of the American Psychological Association, 1966.

Kenny, J., Hentschel, G. and Elpers, K. How Students See Teachers, 1972 (ERIC number ED 077 921).

Morsh, J. E., Burgess, G. and Smith, P. N. Student achievement as a measure of instructor effectiveness. Journal of Educational Psychology, 47:79-88, 1956.

Podlogar, M. and Rosenshine, V. B. The teacher's ability to explain: Evidence on its generalities and correlates. Palo Alto, California: Stanford University, Research Memo No. 10, 1967.

Shaw, J. S. Students evaluate teachers and (better sit down) it works. Nation's Schools, 91 (4), April 1973, 49-53.

Smith, J. P. and Brown, T. J. Relationships among secondary students' evaluations of their courses and teachers and their general attitudes toward school. Paper presented at the American Educational Research Association Annual Meeting, San Francisco, 1976.

Thompson, J. E. Student evaluation of teachers. NASSP Bulletin, October 1974.

Tofelson, N. Selected student variables and perceived teacher effectiveness. Education, 4 (1): 30-35, 1973.

VanNelson, C. and Cossart, E. Does the instructor or the course make the most difference in student evaluation of instruction? Paper presented at the Annual Meeting of the American Eduational Research Association, Washington, D.C., March-April 1975.

Veldman, D. J. and Peck, R. F. The Pupil Observation Survey: Teacher Characteristics from the Students' Viewpoint. Austin, Texas: Texas University, Research and Development Center for Teacher Education, 1967.

Appendix

Teacher Observed _____
Date _____
Subject _____
Period _____

Student Observation of Teachers and Teaching Techniques

The purpose of this teacher observation questionnaire is to help teachers by enabling them to see their strengths and weaknesses through the eyes of their students. Please answer these questions honestly and fairly and keep in mind that this observation questionnaire will in the long run help you, the student, as well as the teacher.

Directions: Please answer each question by circling the number that best describes your teacher. Upon completion of this questionnaire, please follow the directions given to you by your teacher.

1. The teacher is willing to help you when you need it.	5 Always	4	3 Sometimes	2	1 Never
2. The classroom atmosphere created by the teacher is relaxed.	5 Always	4	3 Sometimes	2	1 Never
3. The teacher is respected by you.	5 Always	4	3 Sometimes	2	1 Never
4. The teacher willingly accepts constructive criticism from students.	5 Always	4	3 Sometimes	2	1 Never
5. The teacher values student opinions.	5 Always	4	3 Sometimes	2	1 Never
6. The teacher is fair and impartial in dealing with all students.	5 Always	4	3 Sometimes	2	1 Never
7. The teacher made it clear how grades would be determined.	5 Absolutely clear	4	3 Somewhat clear	2	1 Unclear
8. Tests are returned without too much delay.	5 Always	4	3 Sometimes	2	1 Never
9. The material on tests was covered by the teacher prior to the test.	5 Thoroughly	4	3 Moderately	2	1 Not at all

10. The teacher gives adequate time to complete tests.	5 Always	4	3 Sometimes	2	1 Never
11. The tests measure your understanding of the main ideas of the course.	5 Always	4	3 Sometimes	2	1 Never
12. Tests are explained after they are graded.	5 Always	4	3 Sometimes	2	1 Never
13. Audio-visual aids are on your level of understanding.	5 Under-standable	4	3 Somewhat Difficult	2	1 Difficult
14. If a textbook is used, it provides you with a good understanding of the course.	5 Outstanding in value	4	3 Confusing	2	1 Outdated & irrelevant
15. Textbooks are on your level of understanding.	5 Under-standable	4	3 Somewhat Difficult	2	1 Difficult
16. The teacher is enthusiastic about the subject being taught.	5 Always	4	3 Sometimes	2	1 Never
17. The teacher is patient.	5 Always	4	3 Sometimes	2	1 Never
18. The teacher is cheerful.	5 Always	4	3 Sometimes	2	1 Never
19. Discussions are a meaningful part of this course.	5 Always	4	3 Sometimes	2	1 Never
20. Lectures are a meaningful part of this course.	5 Always	4	3 Sometimes	2	1 Never
21. The teacher stimulates discussions.	5 Always	4	3 Sometimes	2	1 Never
22. The materials of the course are presented in a manner that is easily understood.	5 Always	4	3 Sometimes	2	1 Never
23. The teacher varies manner of instruction.	5 Often	4	3 Sometimes	2	1 Seldom

24. The teacher remains on one subject area until most students understand it.	5 Always	4	3 Sometimes	2	1 Never
25. Class presentation is well organized.	5 Very well	4	3 Moderately organized	2	1 Totally unorganized
26. Class management and discipline is sufficient to avoid disruption of the learning process.	5 Always	4	3 Sometimes	2	1 Never
27. Teacher comes to class well prepared.	5 Always	4	3 Sometimes	2	1 Never
28. The amount of time the teacher expects you to devote to his/her class is:	5 Just right	4	3 Sometimes too much or too little	2	1 Always too much or too little
29. The homework assignments help you understand the course.	5 Always	4	3 Sometimes	2	1 Never
30. Does the teacher have any personal mannerisms that are distracting?	Yes	No	If yes, explain. _____		
31. Would you recommend this course to another student?	5 Definitely	4	3 Perhaps	2	1 No

32. I took this course because:

1. I wanted to take it.
2. I needed it to graduate.
3. I needed it for college.
4. Other - Explain

_____
_____
_____
_____

Thank you for answering each question honestly and completely. If there are any comments you would like to make, please include them below or on the other side of this sheet.